

## Explaining Your Terminology to a Computer\*

Nels E. Olson<sup>1</sup>, David D. Sherertz<sup>1</sup>, Mark S. Erlbaum, MD<sup>1</sup>, Stephanie S. Lipow<sup>1</sup>,  
Octavio Suarez-Munist<sup>1</sup>, Lloyd F. Fuller, PhD<sup>1</sup>, Keith E. Campbell, MD<sup>2</sup>,  
Mark S. Tuttle<sup>1</sup>, Stuart J. Nelson, MD<sup>3</sup>, William G. Cole, PhD<sup>4</sup>

<sup>1</sup>Lexical Technology, Inc., Alameda, CA; <sup>2</sup>Stanford University, Stanford, CA;

<sup>3</sup>Medical College of Georgia, Augusta, GA; <sup>4</sup>Lexical Technology, Inc., Seattle, WA

If you can explain your terminology to a computer, then the maintenance of your terminology may benefit from economies of scale. For example, if one terminology needs an update in response to some change in medical knowledge or practice then it is likely that other terminologies will need to be updated too, and a computer may be able to figure out which maintainers to notify. Further, the software that helps to maintain one terminology may be usable to help maintain others.

But these economies of scale require a common conceptual framework (for humans) and a common infrastructure (for computers). In collaboration with the National Library of Medicine and others, Lexical Technology, Inc. has developed a conceptual framework and some of the requisite infrastructure. Part of the framework is illustrated in **Figure 1**, an adaptation of the semiotic triangle.<sup>1</sup>

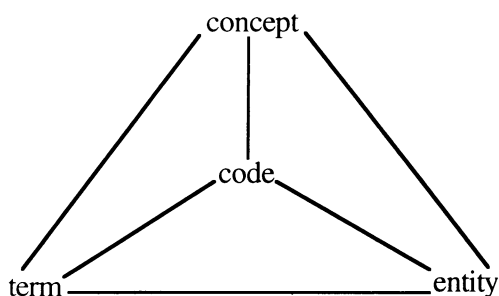


Figure 1 - Units of Meaning.

A *concept* is a unit of thought, a *term* is a unit of language, an *entity* is a unit of reality, and a *code* is a unit of a partition. *One way to describe a terminology is to map it to the units of meaning in Figure 1, and then specify the relationships among those units.* But how does one do this? One process for determining the representation of meaning in a terminology is displayed in **Figure 2**. Beginning at the bottom, we assume that a terminology is represented digitally on some *medium*. This medium has to be readable by some device so that its structure, typically, one or more *files*, becomes explicit. Usually all files share an *encoding*, e.g., EBCDIC or ASCII. Once the encoding is determined, most files have *records*, or entries. In turn, records contain *fields*, and each field contains a (relatively) atomic data element. The next step is to determine how one data element, typically a term, *references*, e.g., "cross references," another data element, typically another term. Does **term A** refer to **term B** by using an abbreviation of **term B**, or does it use **B's** code? Once the method of reference is determined, then the

references need to be interpreted. We think of the references as *propositions*, e.g., that term **A** has terms **B**, **C**, and **D** as children.

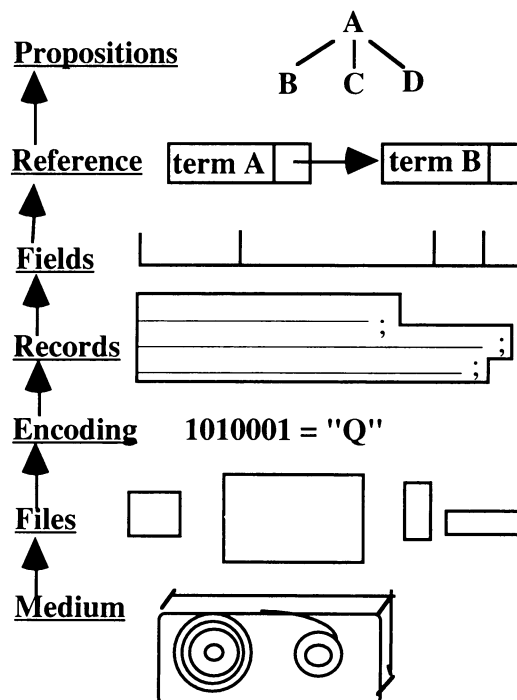


Figure 2 -- The Representation of Meaning

In summary, the process outlined in **Figure 2** produces part of the description of a terminology, namely the terms and codes displayed in **Figure 1**. If the process implied by **Figure 2** were standardized, it could serve as the common infrastructure, not only for description of the remaining units of meaning but also for maintenance of terminologies. While some propositions permit a computer to suggest relationships between one terminology and another, for the foreseeable future a human will be required to confirm any resulting connections. By agreeing on the conceptual framework (**Figure 1**) and a common infrastructure (**Figure 2**), the remaining resources could be focused on those tasks of terminology maintenance that can only be done by humans.

### References

- \* Partially supported by National Library of Medicine contract N01-LM-3-3515.
- 1. Ingenerf, J. Taxonomic Vocabularies in Medicine. *MEDINFO95*, RA Greenes, et al., editors, IMIA 1995, 136-9.